DOCUMENT RESUME

ED 054 195 TM 000 759

AUTHOR Berty, Ernest
TITLE A Guide to the Evaluation of Pilot Programs.
INSTITUTION West Virginia State Dept. of Education, Charleston.
PUB DATE Jan 70
NOTE 19p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Community Characteristics, Data Collection, Educational Change, *Educational Programs, Evaluation Needs, *Evaluation Techniques, *Guides, *Pilot Projects, Post Testing, Pretesting, Program Costs, Program Design, *Program Development, Program Effectiveness, School Environment

ABSTRACT

This guide draws from numerous evaluation models to provide a frame of reference in the development and execution of an evaluation plan for pilot programs. Emphasis is on the development of an evaluation design; construction of instruments or the analysis and interpretation of results is not discussed. A selected bibliography for further in-depth study is included. (AG)

ED054195

A GUIDE TO THE EVALUATION OF PILOT PROGRAMS

January 1970

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Ernest Berty
Office of Research
West Virginia Department of Education
Charleston, West Virginia

VT 000 259

A GUIDE TO THE EVALUATION OF PILOT PROGRAMS

This document is designed for the beginner in practical evaluation who needs a point of departure, especially in dealing with pilot programs or projects. For many reasons, evaluation has suddenly achieved a new prominence. Yet the current state of the art of evaluation leaves many educators somewhat bewildered as to its application. Numerous evaluation models, such as the Tylerian model, CIPP from Ohio University, EPIC from Arizona, Stake's composite-goal model, and others are presently being advocated across the country--each worthy of consideration in its own right. This guide makes no attempt to adhere to any particular model but instead draws from several of them in an attempt to provide a frame of reference for a workable approach. This deliberate avoidance of the use of any particular model is based upon the premise that too much confidence in any one model causes practitioners to become complacent in seeking newer and better approaches.

WHAT IS EVALUATION?

The purpose of an educational pilot program is to determine whether an activity will lead to increased knowledge and skills for those exposed to the program. If those exposed demonstrate significant gains as a result of the program, then it is expected to be incorporated into the overall curriculum on a permanent basis. Obviously, a pilot program enables us to test a particular activity without making a full-scale commitment.

Evaluation is used to ascertain the effectiveness of a pilot program. Evaluation should be regarded as a professional tool which is employed as a means of encouraging program modification and revision as well as either the abandonment of the program altogether or the adoption of the activity as

a permanent part of the educational enterprise. Evaluation in its simplest sense can be considered as the collection of data for the purpose of making decisions.

WHY EVALUATE?

An evaluation plan should not be included in a project proposal solely because it is a legal requirement or mandated by the guidelines. Decision-makers in the school must really believe that evaluation is worth the time and effort required. Yet, we cannot assume that a new and promising approach confirmed on the basis of a good evaluation plan will automatically survive and become a permanent part of the educational enterprise in a particular school system. Careful consideration must be given, prior to any venture into a pilot project, to the past history of the change process in a particular school setting relative to the following questions:

1. What are the mechanisms for change which exist within the system?
2. How prone to change is the system?
3. How will the local decision-makers react to positive results?

Some provision must be anticipated for the adoption or expansion of a worthwhile activity well in advance, probably at the time a decision is made to submit a project proposal. Otherwise, the whole purpose of a pilot program is defeated before the venture begins.

WHEN DO YOU BEGIN TO EVALUATE?

The development of an evaluation plan relative to any pilot program either begins when you first begin to think about the program, or even before. You may be asking yourself how the latter is possible. Sometimes in the execution of a feasibility study or survey to identify and support unmet needs, some very pertinent data will be gathered. Some of this baseline

or benchmark data can be very helpful in determining the direction of the change process, both during the actual implementation of the program as well as in determining the extent to which specified objectives have been reached. These data are very helpful in revealing any condition which may have existed prior to the program activity and which are related to the outcomes.

Thus, evaluation must be an integral part of a pilot program. Evaluation needs to be designed into the program. It should not come as an after-thought to an already existing program, but must become a basic part of the program itself.

CAN AN EVALUATION BE OBJECTIVE?

One of the primary factors which has contributed to the slow growth of formal evaluation as a professional tool is the sensitivity educators have to criticism. Although the "politics" of evaluation is a very interesting subject, it will only be treated briefly here. Evaluation should and must be approached objectively. An evaluator should not set out to prove any particular point of view. The evaluation process should be executed sincerely with no preconceived notions of what the end results should be or must be. Too often educators have so much professional pride and prejudice invested in a program that they are very reluctant to accept any objective evidence which does not support their own convictions. The state of the art of formal evaluation will only begin to improve at an increased rate when we are able to reduce this type of resistance to objectivity to a bare minimum.

WHO SHOULD EVALUATE?

A decision as to whether the pilot program will be evaluated internally,

externally, or through a combination of both approaches is influenced by many factors. For this purpose, internal evaluation is considered to be that which is conducted by the personnel directly responsible for executing the program. Whereas, external evaluation is that which is performed by an independent contractor, such as a university or some other non-profit agency. In any case it is the project director's job to determine who can supply the best information on the effectiveness of the program based upon the availability of fiscal resources, the experience of the personnel, the general complexity of the pilot program, and so forth. A further brief analysis of external evaluation may serve to clarify the ramifications of this decision.

The general "rule of thumb" in estimating the costs of providing for an evaluation by an external agency falls within, but is not limited to, 2 - 5 percent range of the total cost of the project. Some of the pertinent factors which will affect the costs are enumerated below:

1. The general complexity of the evaluation plan.
2. The number of measuring instruments the contractor must develop.
3. The source and experience of the personnel.
4. The agency which has the major responsibility for data collection.
5. The amount of travel involved by the evaluators.
6. The type of final report desired.

In arriving at an agreement with an independent agency, the program administrator should request a letter from the contractor which includes the following:

1. An enumeration of what the evaluator will do.
2. An enumeration of what the pilot program personnel will do.
3. The contractor's charge for the development of any measuring instruments.
4. Specifications of the caliber and experience of the personnel involved.

5. A schedule of payment over a definite period of time, such as:

   a. ¼ when contract is negotiated
   b. ¼ three months later
   c. ¼ six months later
   d. ¼ after delivery of the final report

HOW CAN THE SCOPE AND DIRECTION OF THE EVALUATION BE DETERMINED?

In the development of an evaluation plan there are several early decisions the evaluator must make which will determine to a large extent the direction of the evaluation. For the most part many of these decisions will be judgmental in nature based upon the general complexity of the pilot program. Some of the questions which will have to be answered before decisions can be made include the following:

1. Is this evaluation to be undertaken within a single program or as a comparison between two or more programs?

2. Will this evaluation measure outcomes alone?

3. Or, will the evaluation also attempt to consider the conditions existing prior to the program which may relate to the outcomes as well as the many encounters that occur as the program progresses?

4. What are the variables which will be selected for evaluation as indicated by the objectives of the program?

5. Can the variables referred to in questions 3 and 4 above be stated as measurable objectives?

6. What are the costs involved in implementing each of the alternatives of the decision?

7. Do the project guidelines encourage local directors to include line items in the budget for evaluation?

8. For what audiences will the final evaluation report be prepared?

SOME VARIABLES TO CONSIDER IN A PILOT PROGRAM

I. COMMUNITY CHARACTERISTICS

___ 1. Population
___ 2. Location
___ 3. Occupations
___ 4. Unemployment
___ 5. Welfare
___ 6. Literacy
___ 7. Ethnic Groups
___ 8. Dwellings
___ 9. Delinquency
___ 10. Family Income

II. SCHOOL CHARACTERISTICS

___ 1. Per Capita Expenditures
___ 2. Teachers' Salaries
___ 3. Grade Levels
___ 4. Condition of Facilities
___ 5. Teacher-Pupil Ratio
___ 6. Pupil Enrollment
___ 7. Grouping Practices
___ 8. Curriculum
___ 9. Services Available
___ 10. Achievement Level
___ 11. Teacher Qualifications
___ 12. Teaching Experience
___ 13. Average Age Teachers
___ 14. Male-Female Ratio of Teachers
___ 15. Teacher Turnover
___ 16. Average Daily Attendance
___ 17. Ethnic Groups
___ 18. Male-Female Ratio, Pupils
___ 19. Efforts to Gain Public Acceptance
___ 20. Problems in Gaining Public Acceptance

III. THE PROGRAM

___ 1. Additional Personnel
___ 2. Regular Staff
___ 3. Qualifications
___ 4. Experience
___ 5. Duties
___ 6. Time Commitment
___ 7. In-service Training
___ 8. Male-Female Ratio
___ 9. Provisions of Services
___ 10. Length of Program
___ 11. Hours of Instruction
___ 12. Intervals Between Testing
___ 13. Teacher Meetings
___ 14. Purposes of Meetings
___ 15. Location of Classes
___ 16. Physical Arrangements
___ 17. Grouping of Teachers
___ 18. Grouping of Students
___ 19. Experimental Class
___ 20. Control Class
___ 21. Major Segments of Program
___ 22. Time Devoted Each Activity
___ 23. Pupils Involved
___ 24. Instructional Materials
___ 25. Teacher Activity
___ 26. Aide or Adult-Pupil Ratio
___ 27. Teaching Methods
___ 28. Motivation of Pupils
___ 29. Equipment and Materials Used
___ 30. Instructional Materials Development
___ 31. Parent and/or Community Development
___ 32. Total Cost
___ 33. Broad Categories of Expense
___ 34. Normal Per-Pupil Cost
___ 35. Per-Pupil Cost of Program

IV. EVIDENCE OF CHANGE

___ 1. Changes of Achievement Measures
___ 2. Changes of Method of Instruction
___ 3. Changes in Pupils' Attitudes
___ 4. Changes in Teachers" Attitudes
___ 5. Selection of Control Group
___ 6. Affect by Other Programs
___ 7. Attrition of Participants
___ 8. Pretest and Posttest Samples
___ 9. Characteristics of Experimental Sample
___ 10. Selection of Measuring Instruments
___ 11. Parallel Forms for Pre and Post-testing
___ 12. Conditions When Measures Acquired
___ 13. Qualifications of Assessors or Observers
___ 14. Special Training of Testers
___ 15. Quantify Data
___ 16. Graphically Display Data
___ 17. Basic for Comparison
___ 18. Appropriateness of Statistical Tests
___ 19. Communication of Statistical Conclusions
___ 20. Educational Importance of Conclusions

HOW CAN MEASURABLE OBJECTIVES BE DEVELOPED?

One of the most important steps in the evaluation process is translating broad, vague goals into more specific objectives or outcomes. The process of determining the goals of a pilot program is primarily a rationale and judgmental matter. These judgments may be made by various groups and arrived at by various means (e.g., conference and committee discussion). However, the initial goals which emerge do not usually provide a suitable basis for an evaluation plan, since they are usually expressed in very general terms. Goals expressed in general terms frequently are vague, convey different meanings to different people and, thus, are far removed from the practical operation of appraising. The difficult task is that of translating or expanding the global goals into measurable objectives which can serve as a useful framework for the appraisal.

No attempt will be made herein to give full treatment to the translation of general goals into measurable objectives. Nevertheless, here are some pertinent points to remember in the translation or expansion process:

1. Much professional judgment is involved.
2. Two different groups of professionals will not necessarily construct the same measurable objectives for the same general goal.
3. It is not necessary to have a one-to-one ratio or correspondence between general goals and measurable objectives. More often than not you will have several measurable objectives for each general goal.
4. The translation or expansion procedure is actually a thinking process fortified with a few structural applications.
5. Each measurable objective will probably include the following four component parts:
    a. The target (e.g., student teacher, administrator)

b. The type of behavior (e.g., cognitive, affective, psychomotor).

c. The content area (e.g., subject, mode of instruction, materials).

d. The method of measurement (e.g., standardized test, rating scale, checklist).

HOW CAN EVALUATION ACTIVITIES BE SCHEDULED?

Let's assume at this point in time that the evaluation plan is designed to collect data relative to input variables, program activity variables, and outcome variables, and that the technology does exist for conducting the evaluation or that the time and talent is available to develop the technology. Thus, the next thing we would want to consider is the pacing of the evaluation activity. The work should be reasonably well distributed so that it will not be congregated in such a way that the staff will be overburdened and detract from the quality of work. A practical approach in avoiding a congested schedule may consist of the following three-pronged approach. The first step would entail the enumeration of all of the input, program activity, and outcome variables which will be considered during the evaluation process. The following outline is a brief illustration of this procedure as it applies to a teacher preparation program:

A. Input Variables

1. Publicity

a. Scope
b. Effectiveness of Communication
c. Timing

2. Participants

a. Biological and Professional Data
b. Formal Courses
c. Informal Activities

The second step consists of scheduling the data gathering activity. The following partially-filled version of a table illustrates when the data identified in the first step will be gathered:

	DATA			
	INPUT VARIABLES		PROGRAM VARIABLES	OUTCOME VARIABLES
	A. 1. a-c	A. 2. a-c		
May '69	A_1			
June '69		A_2		
July '69				
August '69				

In the third step the evaluation activities are identified by letters and a brief explanation given about the collection of the data. The following is an example of this step which is consistent with the activity alluded to in the first two steps:

A. The scope, timing, and effectiveness of publicity will be assessed by the means of questionnaires or telephone surveys of a random sample of teachers in the field. In addition, participants will also be questioned about where they heard of the program, how, when, whether publicity was too late or appropriately timed, etc.

Please be reminded at this point that the measuring instruments used by evaluators will include, but not be limited to, inventory schedules, biographical data sheets, interview routines, checklists, opinionaires, and various kinds of psychometric tests.

HOW WILL THE DATA BE COLLECTED?

Since evaluation is the collection of data for decision making, careful consideration must be given to the handling of these data. There are at least two bases of judging the characteristics of a program relative to the data gathered:

1. With respect to standards as reflected by professional judgments, and
2. With respect to relative standards as reflected by characteristics of alternate programs.

The following table is designed to guide in the treatment of the data gathered:

EVALUATION DESIGNS	
DATA WITHIN PROJECT	DATA OUTSIDE PROJECT
1. Project by Comparison with Absolute Standard	1. Project Group Compared with "National" Norm
2. Pretest-Posttest Comparison	2. Change in Project Group Compared with Change in Group During Previous Period
3. Project Group Compared with a Projection	3. Change in Project Group Compared to Change in Concurrent Control Group

HOW CAN PRE AND POSTTESTING BE FACILITATED?

The Solomon Four-Group Design lends itself well to pre and posttesting in short term projects. The design is as follows:

1. Experimental group	Test $_1$	Program	Test $_2$
2. Control Group	Test $_1$		Test $_2$
3. Experimental Group		Program	Test $_2$
4. Control Group			Test $_2$

In this design the experimental group participates in the pilot program and is compared to a group which does not participate. A special effort is made to reduce any effect the pretest itself may be a cause of the pretest-post-test difference. There are four groups involved as illustrated in the table, two experimental and two control groups. The posttest is administered to all four groups; however, only one of the experimental and control groups are pretested. Obviously, a crucial element here is also the assignment to the groups.

HOW CAN THE MEASUREMENT OF PARTICIPANTS BE FACILITATED?

If an evaluator wants to examine the effects of a pilot program it is not necessary for every student in the program to take every item in every test related to the outcomes. Obtaining data on every participant on every item can be a waste of time and effort. To make decisions about programs one needs data on the program, not the individual. If a group is exposed to a program and a set of test items is developed to assess the effects of that program, then the information desired is how well the total group did on the entire set of items. The process known as matrix sampling is a practical approach which facilitates the achievement of this end.

The fundamental idea of matrix sampling is this: every program participant (from a universe of participants) need not respond to every item (from a set of test items) in order to obtain estimates of the mean and variance (standard deviation) of the participants' responses to the set of test items. Matrix sampling involves the simultaneous and random sampling of both participants and test items. The most efficient use of this technique involves different, non-overlapping samples of participants taking non-overlapping samples of test items. That is, one sample of participants take one sample of test items, a second sample of participants takes a second sample of items, and so forth.

Perhaps the best way to describe how to use matrix sampling is to illustrate its use in a hypothetical situation. An evaluator desires to find out how well a particular class is doing. There are 500 students and he has constructed a 50-item test on the curriculum. This is what the evaluator does. He randomly divides the 50-item test into five parts of ten items each. He then randomly administers each group of ten-test items to five groups of 100 students as depicted in the following:

STUDENT SAMPLE (100 each group)	SUBTESTS (10-items each)				
	A	B	C	D	E
I	X				X
II	X	X			
III		X	X		
IV			X	X	
V				X	X

For each of the groups an arithmetic mean and a variance estimate is computed. The arithmetic mean of the five-group means is the best estimate of the total class's average performance. The arithmetic mean of the five group variance estimates is the best estimate of the total class's variance. It is important in this process to have every test item responded to as often as every other item; thus, if necessary, data should be deleted to satisfy this objective.

There is an alternative procedure which can be used, and there are instances in which it may be more desirable. Consider an experimental group of 500 subjects. Instead of using matrix sampling to estimate the means and variance of the 500 subjects on the 50 items, just the mean of each item is estimated. This results in 50 numbers, each number an estimate of the mean of an item for the 500 subjects. The same procedure can be followed for 500 subjects in a control group. Thus, 50 pairs of numbers are obtained, two estimated item means for each of the 50 items. A *t*-test for matched pairs can be performed on these data to examine the hypothesis of no difference between the groups. Since the item data can be useful for diagnostic purposes in the examination of the pilot program, this approach will be valuable for many evaluation situations.

IN RETROSPECT:

As was indicated in the very first paragraph, the suggestions found herein are offered as a frame of reference in the development and execution of an evaluation plan. Its primary emphasis is to assist in the development of an evaluation design. No attempt was made to inform the reader in regard to the development of instruments or the analysis and interpretation of the results. The inclusion of the numerous references in the Selected Bibliography should serve well for further in-depth study of any aspect of evaluation. Also, a deliberate effort was made to avoid advocating the use of any of the current nationwide evaluation models.

Evaluation requires a systematic procedure for marshaling and presenting objective evidence to support any judgment about the effectiveness of a program. Its ultimate justification rests in determining how much is accomplished relative to the program objective, the changes which occurred, and whether or not the changes were the ones which were intended. Nevertheless, it should be reiterated that the potential evaluator should also be alerted to the importance of assessing the process or effort associated with the program. Although the evaluation of the processes has been compared to "the measurement of the number of times a bird flaps his wings without making any attempt to determine how far the bird has flown," an analysis of the process can reveal why a program is not working as well as expected. Locating the cause of the failure can result in modifying the program while it is in progress so it will work. This procedure is a pertinent part of the evaluation plan. It requires a breakdown and assessment of the component parts of the program and the identification of those aspects which contribute or detract from the overall effect of the program.

SELECTED BIBLIOGRAPHY

EVALUATION OF PROGRAMS

Rationale Development

Berlak, H., and Tom, A. Toward rational curriculum decisions in the social studies. St. Louis: Metropolitan St. Louis Social Studies Center, Washington University, 1967. (See pages 10-20)

Broudy, H. S., Smith, B. O., and Burnett, J. R. Democracy and excellence in American secondary education. Chicago: Rand McNally, 1964. (See chapters 1-5)

Bruner, J. S. Toward a theory of instruction. Cambridge: Harvard University Press, 1966.

Hunt, M. P., and Metcalf, L. E. Teaching high school social studies. New York: Harper and Row, 1955. (See chapter 10)

Evaluation Models

Baker, R. L. Curriculum evaluation. Review of Educational Research, 1969, 39, 339-358.

Glass, G. V. The growth of evaluation methodology. Boulder, Colorado: Laboratory of Educational Research, University of Colorado, 1969.

Stake, R. E. The countenance of educational evaluation. Teacher College Record, 1967, 68, 523-540.

Stake, R. E., Tyler, R. W., Gagne, R. M., Scriven, M., and Ahmann, J. S. Perspectives of curriculum evaluation. Chicago: Rand McNally, 1967.

Stufflebeam, D. L. Evaluation as enlightenment for decision-making. Columbus: Evaluation Center, Ohio State University, January 19, 1968.

Suchman, E. A. Evaluative research. New York: Russell Sage Foundation, 1967.

Taylor, P. A., and Maguire, T. O. A theoretical evaluation model. Manitoba Journal of Educational Research, 1966, 1, 12-17.

Tyler, R. W. The functions of measurement in improving instruction. In E. F. Lindquist (Ed.), Educational measurement, Washington: American Council on Education, 1951. Pp. 47-67.

Educational Objectives

Bloom, B. S., and others. Taxonomy of educational objectives: cognitive domain. New York: David McKay, 1956.

Krathwohl, D. R. The taxonomy of educational objectives - use of the cognitive and affective domains. In C. M. Lindvall (Ed.), Defining Educational Objectives. Pittsburgh: University of Pittsburg Press, 1964. Pp. 19-36. Also in N. E. Gronlund (Ed.), Readings in Measurement and Evaluation. New York: The Macmillan Company, 1968. Pp. 18-36.

Krathwahl, D. R., and others. Taxonomy of educational objectives: affective domain. New York: David McKay, 1964.

Mager, R. F. Preparing instructional objectives. Palo Alto, California: Fearon Publishers, 1962.

Woodruff, A. D. A map of classroom conditions required for producing behavorial change in students. Salt Lake City: University of Utah, 1968.

Achievement Measurement

Adkins Wood, D. Test construction: development and interpretation of achievement tests. Columbus, Ohio: Merrill, 1960.

Ahmann, J. S., and Glock, M. D. Evaluating pupil growth: principles of tests and measurement, (3rd ed.) Boston: Allyn and Bacon, 1967. (See chapters 1-7)

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, N.J.: Prentice Hall, 1965.

Educational Testing Service, Multiple-choice questions: a close look. Princeton, N. J.: Educational Testing Service, 1963.

Harris, Chester W. (Ed.) Problems in Measuring Change. Madison, Wis.: University of Wisconsin Press, 1963.

Husek, T. R., and Sirotnik, K. Matrix sampling, UCLA Evaluation Comment, Vol. 1, No. 3, December, 1968.

Lindquist, E. F. (Ed.) Educational measurement. Washington: American Council on Education, 1951.

National Society for the Study of Education Forty-Fifth Yearbook, The measurement of understanding. Chicago: University of Chicago Press, 1946.

Stodola, Q. Making the classroom test, a guide for teachers, (2nd ed.) Princeton, N. J.: Educational Testing Service, 1961.

Thorndike, R. L., and Hagen, E. Measurement and evaluation in psychology and education, (3rd ed.) New York: Wiley, 1969. (See chapters 2, 3, and 4)

Test and Scale Selection and Construction

Anastasi, A. Psychological testing. (3rd ed.) New York: Macmillan, 1968.

Cronbach, L. J. Essentials of psychological testing. (2nd ed.) New York: Harper and Row, 1960.

Edwards, A. L. Techniques of attitude scale construction. New York: Appleton-Century-Crofts, 1957.

Guilford, J. P. Psychometric methods. New York: McGraw-Hill, 1954.

Gulliksen, H. Theory of mental tests. New York: John Wiley and Sons, 1950.

Nunnally, J. C., Jr. Tests and measurements. New York: McGraw-Hill, 1959.

Shaw, M. E., and Wright, J. M. Scales for measurement of attitutdes. New York: McGraw-Hill, 1967.

Torgerson, W. S. Theory and methods of scaling. New York: John Wiley and Sons, Inc., 1958.

Unobtrusive Measures and Observation

Simon, A. and Boyer, E. G. Mirrors for behavior: an anthology of classroom observation instruments. Philadelphia: Research for Better Schools, Inc., 1967.

Webb, E. J. and others. Unobtrusive measures: nonreactive research in the social sciences. Chicago: Rand McNally, 1966.

Research Methodology

Campbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Festinger, L., and Katz, D. Research methods in the behavioral sciences. New York: Holt, Rinehart and Winston, 1953.

Glock, C. Y. (Ed.) Survey research in the social sciences. New York: Russell Sage Foundation, 1967.

Kerlinger, F. N. Foundations of behavioral research. New York: Holt, Rinehart, and Winston, Inc., 1964. (See chapter 3 and part 4)

Oppenheim, A. N. Questionnaire design and attitude measurement. New York: Basic Books, 1966.

Payne, S. L. The art of asking questions. Princeton, New Jersey: Princeton University Press, 1951.

Van Dalen, D. B. Understanding educational research. (2nd ed.) New York: McGraw-Hill, 1967.

Statistical Methodology

Diederich, P. B. Short-cut statistics for teacher-made tests. (2nd ed.) Princeton, N. J.: Educational Testing Service, 1964.

Edwards, A. L. Statistical methods for the behavorial sciences. New York: Holt, Rinehart and Winston, 1964.

Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw-Hill, 1965.

Popham, W. J. Educational statistics. New York: Harper and Row, 1967.

Wert, J. E., Neidt, C. O., and Ahmann, J. S. Statistical methods in educational and psychological research. New York: Appleton-Century-Crofts, 1954.

Resources Materials

Buros, O. K. The mental measurements yearbooks: 1938, 1940, Third (1949), Fourth (1953), Fifth (1960), Sixth (1965). Highland Park, N.J.: Gryphon Press.

Ebel, R. L. (Ed.) Encyclopedia of educational research. (4th ed.) New York: Macmillan Company, 1969.

French, J. W., and Michael, W. B. Standards for educational and psychological tests and manuals. Washington: American Psychological Association, 1966.

Gage, N. L. (Ed.) Handbook on research in teaching. Chicago: Rand McNally, 1963.

Lichey, H. G. (Ed.) Educational evaluation: new roles, new means. Chicago: National Society for the Study of Education, 1969.

Periodicals

Research in Education
(Monthly publication for the Educational Resources Information Center, i.e., ERIC)

Review of Educational Research
(Five issues per year devoted to reviews of research in educational areas)

The EPIE Forum
(Monthly publication of Educational Products Information Exchange which includes information about educational materials)